HRN: HAZE-RELEVANT NETWORK USING MULTI-OBJECT CONSTRAINTS FOR SINGLE IMAGE DEHAZING

Shuying HUANG

School of Software
Tiangong University
Tianjin, 300380, China
e-mail: `shuyinghuang2010@126.com`

Mingyang REN

School of Computer Science and Technology
Tiangong University
Tianjin, 300380, China
e-mail: `renmingyang_9@163.com`

Donglei WU

School of Software and Internet of Things Engineering
Jiangxi University of Finance and Economics
Nanchang, 330000, China
e-mail: `wudonglei92@126.com`

Yong YANG*

School of Computer Science and Technology
Tiangong University
Tianjin, 300380, China
e-mail: `greatyangy@126.com`

Weiguo WAN

School of Software and Internet of Things Engineering
Jiangxi University of Finance and Economics
Nanchang, 330000, China
e-mail: `wanwgplus@163.com`

Yingzhi WEI

School of Computer Science and Technology
Tiangong University
Tianjin, 300380, China
e-mail: `weiyz999@foxmail.com`

Abstract. In recent years, some deep learning dehazing methods based on atmospheric scattering model mostly solve the dehazing results by using depth convolution neural networks (CNNs) to estimate the medium transmission map in the model. However, these methods usually ignored the potential correlation between the transmission map and the atmospheric light in the atmospheric scattering model, which can lead to colour distortion and incomplete dehazing in the dehazing results. To address this problem, this paper first presents a novel Haze-Veil model to increase the correlation between the model parameters by constructing an atmospheric veil term. Then, based on the proposed model, a haze-relevant end-to-end network (HRN) is designed to estimate the parameters of this model and directly output the final clear image. In addition, a cost function is designed by defining multi-object constraint cost functions to further establish the connections between the statistical attributes of the hazy image and the out of HRN. Experiments on benchmark images, which include synthesized and real images, show that HRN effectively removes haze and outperforms most of the existing and state-of-the-art dehazing methods.

Keywords: Single image dehazing, deep learning, convolutional neural network, multi-object constraints

Mathematics Subject Classification 2010: 68U10

1 INTRODUCTION

Haze is a natural atmospheric phenomenon in real-world scenarios. In the presence of haze, the distant objects in the haze become blurred, and the colours of these objects lose fidelity. This is because the irradiance received by the camera from the objects is attenuated. Further, the incoming light of the camera blends with light scattered by droplets, dusts and small particles suspended in the atmosphere. As a consequence, the clarity of captured images can be deteriorated, as shown in Figure 1. Additionally, the performance of many outdoor and automatic driving equipment is also degraded. For these reasons, haze removal is highly desired in

---

* Corresponding author

computer vision applications and has attracted a significant interest in the field of image processing and computer vision in recent years [1, 2].

Single-image dehazing has made significant progress over the years, which can be summarized as three types: image enhancement-based methods, model-based methods and CNNs-based methods. Several image enhancement algorithms have been proposed to remove haze from a single image over the years [3, 4, 5, 6, 7, 8]. For example, the early contrast-based method proposed by Tan [3] assumed that the hazy images have lower contrast compared with clear images, and removed haze by maximizing the contrast of per-patch. This method achieved compelling results, especially in the dense haze regions, but the dehazed images often suffered from halo artifacts and distorted colours due to the lack of physical information. Another contrast-based approach designed by Tarel and Hautière [4] removed haze by assuming that the depth-map must be smoothed except in the region with large depth jumps. Although this method has dramatically improved the edge contour of a hazy image and achieved an effective computation, it is invalid to handle the discontinuous scene depth regions and has obviously distorted the colour. Ancuti and Ancuti [5] first proposed a fusion-based method to enhance the visibility of hazy image. This method significantly improves the visibility of enhanced images, but the over-enhancement problem often occurs when dehazing the dense haze. Recently, Galdran [6] proposed a dehazing algorithm based on multiple-exposure image fusion technology.

a) Input hazy image   b) Dehazed result

Figure 1. Single image dehazing results by our method

This algorithm can effectively remove the haze even in bad weather situations. In addition to above methods, the enhancement-based methods such as the histogram-based method [7], and saturation-based method [8], have also obtained considerable dehazing results. However, the common problem of the above methods is that the degradation mechanism of the hazy image is ignored. Due to the lack of appropriate physical information to constrain the statistical distribution of dehazed image, over-enhancement, colour shift and distortion may emerge. To overcome these problems, the model-based methods based on physical models have attracted much attention in recent years.

It is widely known that a hazy image can be regarded as a convex combination of scene radiance and atmospheric light. The combination coefficient is often called the medium transmission. Therefore, in the past several years, the key issue of image dehazing has been extensively studied by estimating the medium transmission. However, the medium transmission depends on the unknown scene depth information in different positions, which leads to the estimation of transmission and haze removal becoming challenging problems. In general, model-based methods estimated the transmission by prior information and assumptions [9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]. The earliest method of Chavez [9] utilized a subtraction method in the darkest object to address the atmospheric scattering correction to remove haze. This method is effective for hazy images with homogeneous scenes. Fattal [10] proposed a dehazing method based on a generic regularity, where pixels of small image patches in natural images can exhibit a one-dimensional distribution in the RGB colour space. Although the contrast of the recovered image is extremely enhanced by this method, the accuracy of classification in the patches cannot be guaranteed. In addition, it is invalid when the statistical hypotheses fail. The dark channel prior (DCP) discovered by He et al. [11] indicated that at least one-colour channel has some pixels with very low intensity in patches of an image which does not contain the areas of the sky. It was effective in transmission estimation and achieved prominent dehazing results in non-sky regions. However, DCP loses effectiveness for the sky region and is computationally expensive in the procedure of soft-matting [21]. It is worth mentioning that soft-matting could be replaced by a standard median filtering [12], guided image filter [22] and guided joint bilateral filtering [13] to improve the computational efficiency. Meng et al. [14] designed a regularization-based dehazing algorithm that imposed a boundary constraint on the medium transmission. This extension algorithm of DCP effectively improved the dehazing quality of the sky region and obtained considerable achievement for most types of hazy images. Nishino et al. [15] proposed a probabilistic method based on Bayesian theory for dehazing. This method regarded the scene albedo and depth as two statistically independent latent layers and formulated a factorial Markov random field (MRF) of the hazy image to jointly estimate the transmission and depth information. Recently, Berman et al. [16] proposed a non-local haze-line prior in which the colours of a haze-free image are well approximated by a few hundred distinct colour lines, and then they employed the clustering algorithm based on the prior in RGB space to restore the clear image. Huang et al. [17] proposed a haze removal method by adopting the robust sparse representation algorithm to estimate the medium transmission, and designed an adaptive illuminance transfer method to optimize the dehazed results.

Recently, researchers heuristically designed various feature maps and employed probability-based methods to estimate the unknown parameters of the physical model. Tang et al. [18] selected four haze-relevant features to establish the optimal feature combination and applied random forests to estimate transmission. Jiang et al. [19] designed a novel saturation, hue and value colour space-based feature, which

was correlated well with the fog density. The feature was used to learn the refined polynomial regression model for optical depth. Zhu et al. [20] proposed a novel linear model to estimate the scene depth information based on the colour attenuation prior, and used a supervised method to learn the parameters of the model. In many cases, the above methods based on priors or assumptions have achieved impressive dehazing performance. However, the above model-based methods strongly depended on the efficiency of the priors and assumptions, so these methods were often invalid when the real-world hazy images did not meet the assumptions or priors.

For the challenging vision task of dehazing, the human brain can quickly and accurately recognize the density and area of fog or haze without any extra information. Inspired by this human behavior, the bio-inspired technology of convolutional neural networks (CNNs) has been successfully used in single image dehazing works [23, 24, 25, 26]. Cai et al. [23] proposed a deep-learning network that was an end-to-end system for medium transmission estimation. This method proposed a creative nonlinear activation function BReLU to improve the quality of the dehazed image. Li et al. [24] proposed cascaded CNNs to jointly estimate the medium transmission and the global atmospheric light under two subnetworks. Ren et al. [25] designed multiscale deep CNNs with coarse and refined networks to extract features, and automatically learned the mapping of hazy images and the corresponding transmission map. Another CNNs-based method [26] tried to bypass the estimation step and merge the parameters of an atmospheric scattering model into one parameter called $K(x)$ for dehazing. Recently, Zhang et al. [27] estimated the transmission map by designing a new edge-preserving densely connected structure and the generative adversarial networks (GAN) framework was employed to evaluate whether the results are real or fake. Ren et al. [28] constructed a CNN by adopting a novel fusion-based strategy, and this network could remove haze without estimating the parameters of any physical model. Wang et al. [29] proposed an algorithm which is based on atmospheric illumination prior to remove haze. The atmospheric illumination prior indicates that the haze mainly affects the luminance channel in YCrCb colour space. The problem of brightness and ambiguity in haze image could be well-handled by using the prior. Chen et al. [30] proposed an end-to-end gated context aggregation network to directly restore the final haze-free image. The latest smoothed dilation technique has been employed in this method to effectively remove the gridding artifacts. In addition, a gated sub-network was designed to fuse the features from different levels to help improve the dehazed result. Song et al. [31] designed a ranking CNNs to extend the structure of CNN, so that the statistical and structural properties of hazy images can be captured at the same time. Yeh et al. [32] proposed a deep learning framework based on multi-scale residual learning (MSRL) and image decomposition for single image defogging. Yin et al. [33] created a colour-transfer image dehazing model and designed a CNN based deep framework based on this model to realize the hazy removal. Dong et al. [34] proposed a multi-scale boosted dehazing network based on boosting and error feedback principles to solve the dehazing problem. Shao et al. [35] proposed a domain adaptation paradigm to address the issue

that most existing dehazing methods are less able to generalize well to real hazy images. Further, Wang et al. [36] proposed a variational single nighttime image haze removal method, which adopts a new structure-aware variational framework to simultaneously estimate the inverted scene radiance and the transmission in the gray component. Yang et al. [37] proposed a progressive image dehazing network, which realizes haze removal step by step by constructing the preliminary and fine dehazing modules.

Zhang and Patel [38] proposed an end-to-end densely connected pyramid network by embedding the atmospheric scattering model directly into the network. Li et al. [39] proposed an unsupervised and untrained neural network for image dehazing, which uses three joint sub-networks to divide the hazy image into three potential layers, namely scene radiance layer, transmission map layer and atmospheric light layer. Li et al. [40] designed an encoder and decoder architecture, which is based on the conditional generative adversarial network (cGAN) to achieve the end-to-end image haze removal. To generate realistic dehazed results, the basic cGAN formulation was modified by introducing the VGG features and adding an L1-regularized gradient prior. In general, because the complex nonlinear mapping from the input to the target can be sufficiently learned by fitting the massive training data, increasingly progress has been achieved by these CNNs-based dehazing methods. Wu et al. [41] propose a new contrast regularization method based on contrast learning, which utilizes information from blurred images and clear images as negative and positive samples, respectively. However, many CNNs-based methods do not fully consider the correlation between parameters in the haze degradation model when designing networks. This problem may result in the features extracted by CNNs being inadequate and inappropriate, and it will further lead to an undesirable dehazing results in real-world images [42].

In this paper, to better explore the physical information of haze by means of CNNs, we build a novel Haze-Veil model by modifying the regular atmospheric scattering model to describe the degradation mechanism of a hazy image. Then, a haze-relevant end-to-end network (HRN) based on the proposed Haze-Veil model is designed to achieve the task of dehazing. HRN takes a hazy image as input, and directly outputs the haze-free image without any manual parameter assignment. In addition, to make HRN more closely related to the physical mechanism of haze, a new cost function is also designed to constrain the learning of HRN. Extensive experiments on benchmark images prove that the proposed HRN obtains better performance over some the state-of-the-art methods. The main contributions of this paper include:

1. A new Haze-Veil model based on the regular atmospheric scattering model is built to describe the degradation mechanism of a hazy image. In the proposed model, the parameter, i.e., atmospheric light is no longer estimated solely but is replaced by estimating a more appropriate parameter called atmospheric veil. The new Haze-Veil model enhances the correlation of model parameters without changing the physical mechanism of the scattering model.

2. A novel and effective dehazing network called HRN is proposed in this paper. HRN jointly learns medium transmission and atmospheric veil, and directly outputs the haze-free image. Further, HRN can realize information interaction between the atmospheric veil and medium transmission, so as to achieve more accurate learning.

3. We propose a new multi-object constraints cost function by defining several constraints items, which are used to generate a constraint relationship between the medium transmission and atmospheric veil. This multi-object constraints cost function can effectively improve the quality of the dehazed results.

The remainder of this paper is organized as follows. In Section 2, the background knowledge is introduced briefly. In Section 3, the details of our proposed dehazing algorithm are elaborated, and the designed motivations of our work are analyzed. The experimental results and comparison are given in Section 4. Finally, the conclusion is summarized in Section 5.

2 ATMOSPHERIC SCATTERING MODEL

The widely used haze model is called the atmospheric scattering model [11], which is written as:

$$\boldsymbol{I}(x) = \boldsymbol{J}(x)t(x) + \boldsymbol{A}(1 - t(x)), \tag{1}$$

where $\boldsymbol{I}(x)$ is the observed hazy image, $\boldsymbol{J}(x)$ is the scene radiance, and $\boldsymbol{A}$ is the atmospheric light, which describes the global light intensity of the scene. x represents the pixel location. $t(x)$ is the medium transmission range of (0,1), which describes the portion of light that directly reaches the camera and is not scattered. $t(x)$ is correlated with the scene depth and can be expressed as follows:

$$t(x) = e^{-\beta d(x)}, \tag{2}$$

where $d(x)$ is the scene depth and β is the atmosphere scattering coefficient. Equation (2) suggests that the medium transmission attenuates exponentially with the scene depth $d(x)$. The first term $\boldsymbol{J}(x)t(x)$ in Equation (1) is called direct attenuation [1], which describes the decay of $\boldsymbol{J}(x)$ in the medium. The second term $\boldsymbol{A}(1 - t(x))$ is called airlight [1], which results in the scattered light and generates a white atmospheric veil in the degraded image.

The image dehazing could be regarded as recovering the scene radiance $\boldsymbol{J}(x)$ from a haze image $\boldsymbol{I}(x)$ based on Equation (1). This requires that the unknown parameters of the medium transmission $t(x)$ and atmospheric light $\boldsymbol{A}$ are appropriately estimated. Once $t(x)$ and $\boldsymbol{A}$ are solved, the haze-free image can be recovered by:

$$\boldsymbol{J}(x) = \frac{\boldsymbol{I}(x) - \boldsymbol{A}}{\max(t(x), \delta)} + \boldsymbol{A}, \tag{3}$$

where δ is a small constant for avoiding the zero divisor.
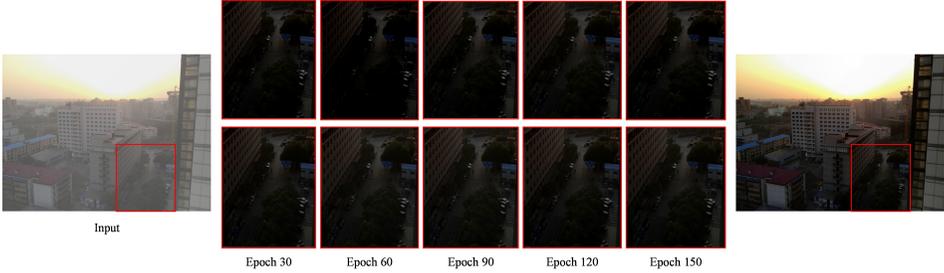
Figure 2. The dehazed results of the different epoch. The images in first row are the visualized results by separately estimating the A and $t(x)$. The images in second row are the visualized results by using our proposed method.

3 THE PROPOSED METHOD

In this section, the proposed method is explained mainly in three aspects. We first proposed the Haze-Veil model and analyze its performance. Then, the HRN which is based on the Haze-Veil model is presented, and the design motivations of the network structure are discussed. Finally, we introduce the proposed cost function and analyze the effectiveness and significance of it.

3.1 Haze-Veil Model

Since we aim to design end-to-end CNNs, it means that all the parameters in the atmospheric scattering model must be learned in the network. According to the atmospheric scattering model introduced in Section 2, the unknown parameters include the haze-free image $J(x)$, atmospheric light A and medium transmission $t(x)$. An intuitive resolution is that the designed CNNs could recover $J(x)$ by jointly learning A and $t(x)$. However, we find that the dehazed images are always too dark or too bright, especially when the hazy image has a bad illuminance condition. According to the analysis of Equation (1), it can be inferred that A and $t(x)$ are two independent variables, that is, there is no correlation between A and $t(x)$. In other words, the different values of A have no effect on $t(x)$, so it will result in different dehazed results in the end. Furthermore, the experiments prove that the network is hard to obtain the satisfactory results when it trains the independent parameters of A and $t(x)$ at the same time. As seen from the first row in Figure 2, when we separately train the A and $t(x)$ at the same time, it is hard to obtain desired dehazed results with the increase of epoch. To overcome this problem, we will not separately estimate A and $t(x)$ anymore but merge the factors in the second term of Equation (1) into one factor to enhance the correlation of A and $t(x)$. The merged term is defined as:

$$V(x) = A(1 - t(x)). \tag{4}$$
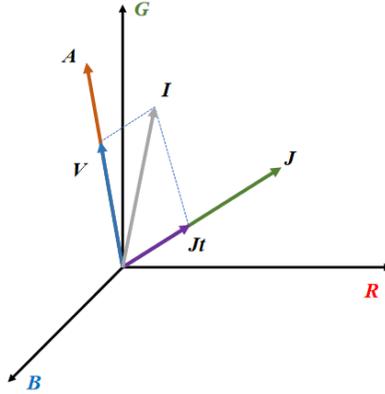
Figure 3. The geometrical relationship of the parameters of Haze-Veil model in RGB space. $\boldsymbol{I}$ is the hazy image, $\boldsymbol{J}$ is the haze-free image, V is the medium transmission, and $\boldsymbol{A}$ is the atmospheric light. $\boldsymbol{V}$ is the atmospheric veil.

The atmospheric veil $\boldsymbol{V}(x)$ includes the information of the atmospheric light $\boldsymbol{A}$ and the medium transmission $t(x)$. In this way, we do not train and estimate $\boldsymbol{A}$ anymore, but estimate the more appropriate $\boldsymbol{V}(x)$, which not only includes information of $\boldsymbol{A}$ but is also relevant to $t(x)$. Therefore, according to Equation (1) and Equation (4), the modified atmospheric scattering model can be written as follows:

$$\boldsymbol{I}(x) = \boldsymbol{J}(x)t(x) + \boldsymbol{V}(x). \tag{5}$$

Equation (5) is called Haze-Veil model, and $\boldsymbol{I}(x)$ and $\boldsymbol{J}(x)$ still represent the hazy image and scene radiance, respectively. $t(x)$ is the medium transmission. Physically, the intuitive meanings of $\boldsymbol{V}(x)$ could be explained as the haze density. In the map of $\boldsymbol{V}(x)$, the regions with heavy haze have high intensity, and the regions with thin haze have low intensity. In fact, it can be considered that there is a negative correlation between $t(x)$ and $\boldsymbol{V}(x)$. Geometrically, the vector $\boldsymbol{V}$ is parameterized by $(1-t)$ and oriented by $\boldsymbol{A}$. $\boldsymbol{V}$ and $\boldsymbol{A}$ are two collinear vectors, and the correlation is illustrated in Figure 3. As seen, $\boldsymbol{J}$ is attenuated and shifted by t and $\boldsymbol{V}$, so a clear image is degraded to a hazy image $\boldsymbol{I}$. By the above analysis, the task of dehazing is changed to recovering $\boldsymbol{J}(x)$ via estimating $t(x)$ and $\boldsymbol{V}(x)$ from Equation (5). Because we no longer detachedly estimate the $t(x)$ and $\boldsymbol{A}$, the recovered model is simpler than before and the network training becomes easier and is more explainable.

3.2 HRN: The Haze-Relevant End-to-End Network

Based on the proposed Haze-Veil model, we design the HRN for removing haze. We first aim to design a network to jointly estimate $t(x)$ and $\boldsymbol{V}(x)$. Then, the dehazed image can be obtained by computing the proposed Haze-Veil model. Therefore, the
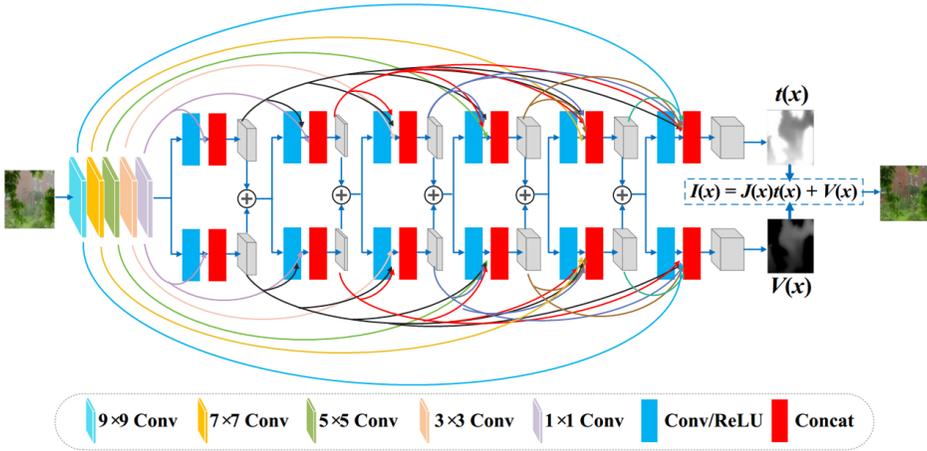
Figure 4. The framework of the proposed HRN structure. Different colour blocks mean multi-scale operations.

proposed HRN is composed of two submodules: the parameters estimation module which is used to extract the features and estimate the $t(x)$ and $\boldsymbol{V}(x)$, and the image recovering module which is used to output the haze-free image according to Equation (5). The proposed HRN achieves the end-to-end operation without any parameter assignment in the whole dehazing procedure. The network framework of HRN is illustrated in Figure 4.

3.2.1 Parameters Estimation Module

Multiscale feature extraction has been proven to be effective and successful in single image dehazing in recent years [23, 25]. It is widely accepted that the convolutional operation in the low level could learn the basic features, such as edges and texture [43]. For the task of image dehazing, these basic features are very important for recognizing the regions with a depth change. Therefore, to fully extract the basic features for both the medium transmission $t(x)$ and atmospheric veil $\boldsymbol{V}(x)$, the multiscale convolutional operations are set in the first shared hidden layer, which is used to extract appropriate features from the input haze image. The sizes of multiscale convolution filters are 1×1, 3×3, 5×5, 7×7, 9×9, and the numbers of these five different scale filters are all set to 4.

(The particular parameters set for each layer can be referred to Table 1 in Section 4). To jointly estimate the $t(x)$ and $\boldsymbol{V}(x)$, we set two parallel subnetworks behind the shared hidden layer to effectively use the multiple features learned by multiscale convolutional filters. The upper branch is for the $t(x)$ estimation and
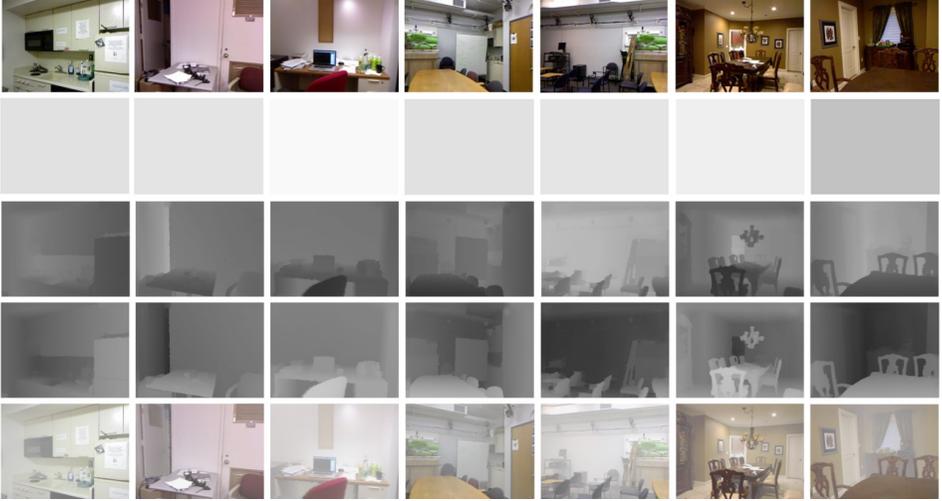
Figure 5. The different feature maps of training datasets. The first row are the clear images $\boldsymbol{J}$. The second row is the random set atmospheric light $\boldsymbol{A}$. The third row is the synthetic atmospheric veil $\boldsymbol{V}$. The fourth row is the synthetic medium transmission t. The fifth row is the synthetic hazy images $\boldsymbol{I}$. In this paper, $\boldsymbol{I}$ is the input of HRN, t and $\boldsymbol{V}$ are ground-truth for training.

the lower branch is for the $\boldsymbol{V}(x)$ estimation. These two subnetworks exactly have the same structure and include seven hidden layers for feature extraction in each subnetwork. To ensure that the information is better flowing along each branch and to speed up convergence of the network, the densely connected structure is embedded into HRN. It leads the information of each layer flowing to the following layers in a feed-forward fashion. In this way, the information of each layer has the direct effect on the final estimation. Further, because the correlation exists in $t(x)$ and $\boldsymbol{V}(x)$, we infer that the correlation should also exist in the features of hidden layers of $t(x)$ and $\boldsymbol{V}(x)$. Thus, to interflow the information of $t(x)$ and $\boldsymbol{V}(x)$, the information interaction structure is designed between the upper and lower branch by conducting the summation operation. Specifically, we sum the feature maps of the same layer in the upper and lower branch. At the end of each subnetwork, the $t(x)$ and $\boldsymbol{V}(x)$ are obtained by the 1×1 convolution operation.

3.2.2 Haze Removal Module

Once the parameters of $t(x)$ and $\boldsymbol{V}(x)$ are estimated by the parameter estimation module, the haze-free image could be recovered via solving Equation (5) and the formula is re-written as follows:

$$\boldsymbol{J}(x) = \frac{\boldsymbol{I}(x) - \boldsymbol{V}(x)}{\max(t(x), \delta)}, \tag{6}$$

where $\boldsymbol{I}(x)$ is the input hazy image, δ is a small constant for avoiding the zero divisor (δ is set to 0.05 in this paper), and $\boldsymbol{J}(x)$ is the output dehazed image. Therefore, the whole inference process of HRN in Figure 5 can be concluded as follows:

Step 1: Hazy image $\boldsymbol{I}(x)$ is input into the HRN, and the multiscale features are extracted by the shared hidden layers.

Step 2: The shared multiscale features are input into the two subnetworks of HRN for further feature extraction.

Step 3: After several times of features extraction in each subnetwork, and the information interaction between the two subnetworks, $t(x)$ and $\boldsymbol{V}(x)$ can be estimated by an appropriate training method, respectively. (The details of training will be described in Section 4).

Step 4: By inputting $t(x)$ and $\boldsymbol{V}(x)$ into Equation (6), the final recovered haze-free image $\boldsymbol{J}(x)$ can be computed.

To further generate the constraints on the correlation parameters $t(x)$ and $\boldsymbol{V}(x)$ and accelerate the convergence of the HRN, an appropriate cost function is essential. Thus, we propose a multi-object constraints cost function to assist the HRN to obtain the better results, and it will be elaborated in the next subsection.

3.3 Multi-Object Constraints Cost Function

In the field of learning-based single image haze removal, the mean squared error (MSE) cost function is often adopted to optimize the network and has achieved outstanding performance. However, because the relevant parameters of $t(x)$ and $\boldsymbol{V}(x)$ are jointly learned by HRN in this paper, the effect of the correlation between $t(x)$ and $\boldsymbol{V}(x)$ cannot be fully used if there is only MSE in the cost function. To further enhance the information interaction between $t(x)$ and $\boldsymbol{V}(x)$, we intend to construct a constraint between the $t(x)$ and $\boldsymbol{V}(x)$ by adding the multi-object constraints cost function about $t(x)$ and $\boldsymbol{V}(x)$ on the regular MSE cost function. According to Equation (4), $t(x)$ can be expressed by $\boldsymbol{V}(x)$ as follows:

$$t = 1 - \frac{V}{A}. \tag{7}$$

According to Equation (4) and Equation (7), $t(x)$ and $\boldsymbol{V}(x)$ can be expressed by each other. Hence, to take full advantage of the correlation, we can employ one's label to constrain the other's output. Therefore, we define three constraints items to yield the constraints in $t(x)$ and $\boldsymbol{V}(x)$, and the constraints items are formulated as follows:

$$L_{C\_t} = \frac{1}{WH} \sum_{w=1}^{W} \sum_{h=1}^{H} \left\| t_{GT} - (1 - \frac{V_{output}}{A_{GT}}) \right\|_2, \tag{8}$$

$$L_{C\_V} = \frac{1}{WH} \sum_{w=1}^{W} \sum_{h=1}^{H} \left\| V_{GT} - (1 - t_{output}) \times A_{GT} \right\|_2, \tag{9}$$

$$L_{C\_Vt} = \frac{1}{WH} \sum_{w=1}^{W} \sum_{h=1}^{H} \left\| V_{output} - (1 - t_{output}) \times A_{GT} \right\|_2, \tag{10}$$

where V_{output} and t_{output} are the outputs of HRN. V_{GT}, t_{GT} and A_{GT} are the ground-truth of $\boldsymbol{V}(x)$, $t(x)$ and atmospheric light $\boldsymbol{A}$, respectively. W and H are the width and height of the input hazy images, respectively. WH is the dimension of the input hazy image. Here, the L_{C_t} and the L_{C_V} are the constraints items designed for $t(x)$ and $\boldsymbol{V}(x)$, respectively. Thus, Equation (8) and Equation (9) mean one of the outputs is constrained by the other's label. Further, L_{C_Vt} is designed for constraining the HRN's output of $t(x)$ and $\boldsymbol{V}(x)$ by each other. In addition, the MSE loss functions are also combined to jointly optimize each subnetwork and are formulated as follows:

$$L_{MSE\_t} = \frac{1}{WH} \sum_{w=1}^{W} \sum_{h=1}^{H} \left\| t_{output} - t_{GT} \right\|_2, \tag{11}$$

$$L_{MSE\_V} = \frac{1}{WH} \sum_{w=1}^{W} \sum_{h=1}^{H} \left\| V_{output} - V_{GT} \right\|_2, \tag{12}$$

where the L_{MSE_t} and L_{MSE_V} are the MSE loss functions of the output of $t(x)$ and $\boldsymbol{V}(x)$, respectively. Finally, by combining MSE loss and constraint loss, the final cost function is defined as:

$$L = L_{C\_t} + L_{C\_V} + L_{C\_V\_t} + L_{MSE\_t} + L_{MSE\_V}. \tag{13}$$

As analyzed before, the final loss function achieves information interaction between $t(x)$ and $\boldsymbol{V}(x)$ during the process of learning.

4 EXPERIMENTS RESULTS

In this section, we demonstrate the effectiveness of the proposed method by extensive experiments on synthetic and real datasets. First, we illustrate the datasets used in this paper. Then, the parameter settings and training details are described. Finally, we conduct subjective comparisons and objective quality assessments between the proposed method and the recent state-of-the-art methods.

4.1 Datasets

It is widely believed that one of the challenges existing in learning-based image dehazing is how to collect a large number of training data. Recently, this problem has been solved by creating the training and test datasets using the public depth dataset of NYU-Depth [44], which includes indoor clear images $\boldsymbol{J}$ and the corresponding depth maps $d(x)$. These two terms are essential for synthesizing the hazy images. The specific process is as follows. First, according to Equation (2), as referenced in [24, 25, 26, 27], we randomly set 10 scattering coefficients $\beta \in [0.5, 1.2]$, and the medium transmission $t(x)$ is synthesized by $d(x)$ and β through Equation (2). Then, we randomly produce 10 atmospheric light values $\boldsymbol{A} \in [0.7, 1]$, and the atmospheric veils $\boldsymbol{V}(x)$ are generated by $\boldsymbol{A}$ and $t(x)$ through Equation (4). Finally, the hazy images are obtained by combining the transmissions $t(x)$, the atmospheric veils $\boldsymbol{V}(x)$ and the clear images $\boldsymbol{J}$ through Equation (5). In HRN, $\boldsymbol{I}$ is the input hazy image, $t(x)$ and $\boldsymbol{V}(x)$ are the labels for training. There are several samples of the dataset as shown in Figure 5. The images in the first row are the haze-free ground-truth images $\boldsymbol{J}$, which are selected from the dataset of the NYU-Depth. The images in the second row are the atmospheric light values $\boldsymbol{A}$, which are randomly set by us. The images in the third row indicate the atmospheric veil maps $\boldsymbol{V}(x)$, which are synthesized through Equation (4), and the images in the fourth row are the synthesized medium transmission maps $t(x)$, which are synthesized through Equation (2). In the last row, the synthetic hazy images are obtained through Equation (5). In this way, the synthesized input images and the corresponding labels are obtained.

It is noted that the heavy haze has a large value of β and that the thin haze has a small value of β. A bright hazy image has a large value of $\boldsymbol{A}$, and a dark hazy image has a small value of $\boldsymbol{A}$. Hence, there are various types of hazy images in our training and test datasets. To further enhance and increase the training datasets, the flip operations for up, down, left and right are conducted on hazy images and the corresponding depth maps. As a result, 40 hazy images can be produced by using a clear image and the corresponding depth map. Two thousand clear images in the NYU-Depth datasets are selected in this paper, so we obtain 80 000 training data in total. In this way, we can effectively improve the robustness of the algorithm and avoid overfitting for training. For the test dataset, we directly use the dataset called RESIDE [45, 46], which is a large-scale benchmark dataset consisting of both synthetic and rea-world hazy images. To quantitatively evaluate our methods, 200 indoor and 200 outdoor synthetic test images are employed to conduct comparison experiments. To further prove the validity of the proposed method, some real hazy images are also used to conduct comparison experiments in this section.

4.2 Parameter Settings and Implementation

Since there are two parallel and same subnetworks in HRN for estimating the medium transmission $t(x)$ and atmospheric veil $\boldsymbol{V}(x)$, we consider that these two

features have almost the same significance for dehazing. Thus, we set the identical parameters in each subnetwork. The parameters of each subnetwork are shown in detail in Table 1. It should be noted that the Concatenation in Table 1 denotes the concatenate operation of the densely structure, and the Summation in Table 1 denotes the information interaction structure between the two subnetworks.

Layers	Operation	Input Size	Filter Number	Filter Size
Shared Layer	Convolution/ReLU	$224 \times 224 \times 3$	4	1×1
			4	3×3
			4	5×5
			4	7×7
			4	9×9
	Concatenation			
Hidden Layer 1	Convolution/ReLU	$224 \times 224 \times 20$	4	3×3
	Concatenation			
	Summation			
Hidden Layer 2	Convolution/ReLU	$224 \times 224 \times 24$	4	3×3
	Summation			
	Summation			
Hidden Layer 3	Convolution/ReLU	$224 \times 224 \times 8$	8	3×3
	Concatenation			
	Summation			
Hidden Layer 4	Convolution/ReLU	$224 \times 224 \times 16$	16	3×3
	Concatenation			
	Summation			
Hidden Layer 5	Convolution/ReLU	$224 \times 224 \times 32$	32	3×3
	Concatenation			
	Summation			
Hidden Layer 6	Convolution/ReLU	$224 \times 224 \times 64$	64	3×3
Output Layer	Convolution/ReLU	$224 \times 224 \times 128$	128	1×1

Table 1. The parameters setting of each subnetwork

The proposed HRN is trained on a NVIDIA 1080Ti GPU using the TensorFlow framework. During the training process, Gaussian random variables (the mean value is 0 and standard deviation is 1) are used to initialize the filters' weights of each layer. The ReLU method is used to activate the neurons. The ADAM [47] optimizer is applied for decreasing the loss. We resize all inputs and labels to 224×224 and use the batch size of 8 for training. The initial learning rate is set to 0.0001, and we decrease it by testing the loss value and PSNR of the validation dataset, so the training of the network is jointly supervised by two indexes in this paper. HRN takes approximately 150 epochs to converge.

4.3 Comparisons on Synthetic Datasets

The synthetic datasets consist of 200 indoor synthetic images and 200 outdoor synthetic images. We evaluate these various methods by SSIM (structure similarity) [48] and PSNR on the synthetic datasets. The higher value indicates that the dehazed image is more similar to the haze-free ground-truth for each of the criteria. We select five different popular dehazing methods for comparison which include: The model-based method is the dark channel prior method (DCP) proposed by He et al. [11], which is commonly used as the baseline method. The CNNs-based methods are the DehazeNet proposed by Cai et al. [23], the DCPDN proposed by Zhang et al. [42], the YOLY proposed by Li et al. [39], the multiscale convolutional neural networks (MSCNN) which is proposed by Ren et al. [25], and the SLA proposed by Liang et al. [49]. To verify the effectiveness of the proposed method, we conduct the comparison experiments on the synthesized test dataset. Five synthesized indoor image samples, which are named Indoor 1–Indoor 5, are shown in the Figure 6. Images in the first row denote the hazy images of the indoor synthetic dataset, the corresponding dehazed results of the compared dehazing methods are shown in the second to sixth rows, and the dehazed results of the proposed method (HRN) are shown in the seventh row. Because the dataset is synthesized, the ground-truth clear images are available and are shown in the last row. For better observation, this paper marks the typical areas in the images with red boxes. As revealed by Figure 6, the dehazed results of MSCNN and YOLY have severe colour distortion, e.g., the door colour of Indoor 2 is unrealistic, and the background colour of Indoor 5 is distorted. The dehazed results of DCP and DehazeNet obtained a realistic colour, but due to the inaccurate estimation of transmission maps, there are still hazes remaining in some areas, such as the blurry door of Indoor 2, the region near the mirror of Indoor 3, and the region near the television of Indoor 4. The results of DCPDN have the problem of over-enhancement. Compared to the results of other dehazing methods, the advantages of our dehazed results are that the problems of colour distortion and overestimation of haze are effectively avoided, so the dehazed images yielded by the proposed method are closer to the ground-truth than other results. As seen, our results in the seventh row are very close to the ground-truth in the eleventh row both in colour and the clarity.

To further testify the proposed methods, we also display five synthesized outdoor samples named Outdoor 1–Outdoor 5, as shown in Figure 7, to show the dehazed results of outdoor images. The images in the first row denote the synthesized outdoor hazy images, the corresponding clear ground-truth images are shown in the last row, the second row to the sixth row show the results of the compared dehazing methods, and the images in the seventh row are our results. It can be observed from Figure 7, due to the inaccurate estimation of transmission maps, the problem of the unrealistic colour cast still exists in the methods of MSCNN, e.g., the tone of the sky city is too dark green in the Outdoor 2. In addition, the results of DCP and DehazeNet are too dark since the medium transmission of them is overestimated, e.g., the contours of buildings have disappeared in the Outdoor 4. Furthermore, Although DCPDN

can remove the hazy in the image, the sky area in the dehazing results will have the problems of over-enhancement. From the dehazing results obtained by YOLY, it can be seen that the hazy in the distant areas is not completely removed, and some areas in the foreground are darkened, causing some information to be hidden, such as the areas in the red boxes in the Outdoor 1 and 3. In contrast, our method removes haze clearly and restores the realistic colours visually. As seen from the results of our method, the colours of the sky region of the Outdoor 1 and the Outdoor 5 are closer to the ground-truth than other results, i.e., the colours of the images in the tenth row are very close to the ground-truth.

To further evaluate the performance of the proposed method, the quantitative comparison is conducted by testing the average PSNR and SSIM of the different methods on the test datasets. As shown in Table 2, the red numbers present the best values, and the blue numbers present the second-best values. From Table 2, it can be observed that the PSNR value obtained by our method for indoor image dehazing is higher than that of other comparison methods, and the obtained SSIM value ranks third; the PSNR value of our method for outdoor image dehazing ranks second in all the results, but the SSIM value is higher than that of other comparison methods.

Datasets	Evaluation	DCP	DehazeNet	MSCNN	DCPDN	YOLY	SLA	Ours
Indoor	PSNR (dB)	22.15	20.98	20.14	21.42	21.18	20.49	22.40
Images	SSIM	0.753	0.779	0.701	0.876	0.752	0.858	0.805
Outdoor	PSNR (dB)	21.16	26.83	23.92	18.97	24.33	24.33	26.47
Images	SSIM	0.853	0.8703	0.929	0.875	0.887	0.932	0.932

Table 2. The average PSNR and SSIM of the comparison dehazing methods on indoor and outdoor synthetic datasets

4.4 Comparisons on Real Images

To further demonstrate the effectiveness of the proposed method, we conduct the comparison experiments on a series of real-world hazy images, and seven commonly used and challenging images, which are named Railway, Pumpkin, House, City and Girls are selected to display in Figure 8.

As seen from Figure 8, the comparison experiments of the proposed methods and the previous five dehazing algorithms are displayed in each row. Images in the first row are the hazy input image, the results of previous dehazed methods are shown in the second row to the sixth row, and the results generated by HRN are shown in the seventh row. Since there is no ground-truth for these real samples, we analyze the experimental results subjectively. As shown in the second and fourth rows, DCP and MSCNN can remove most of the haze, but the dehazed results suffer from the over-enhancement effects, such as the distance region in the Pumpkin and the plant areas in the House. Moreover, the results of MSCNN also suffer
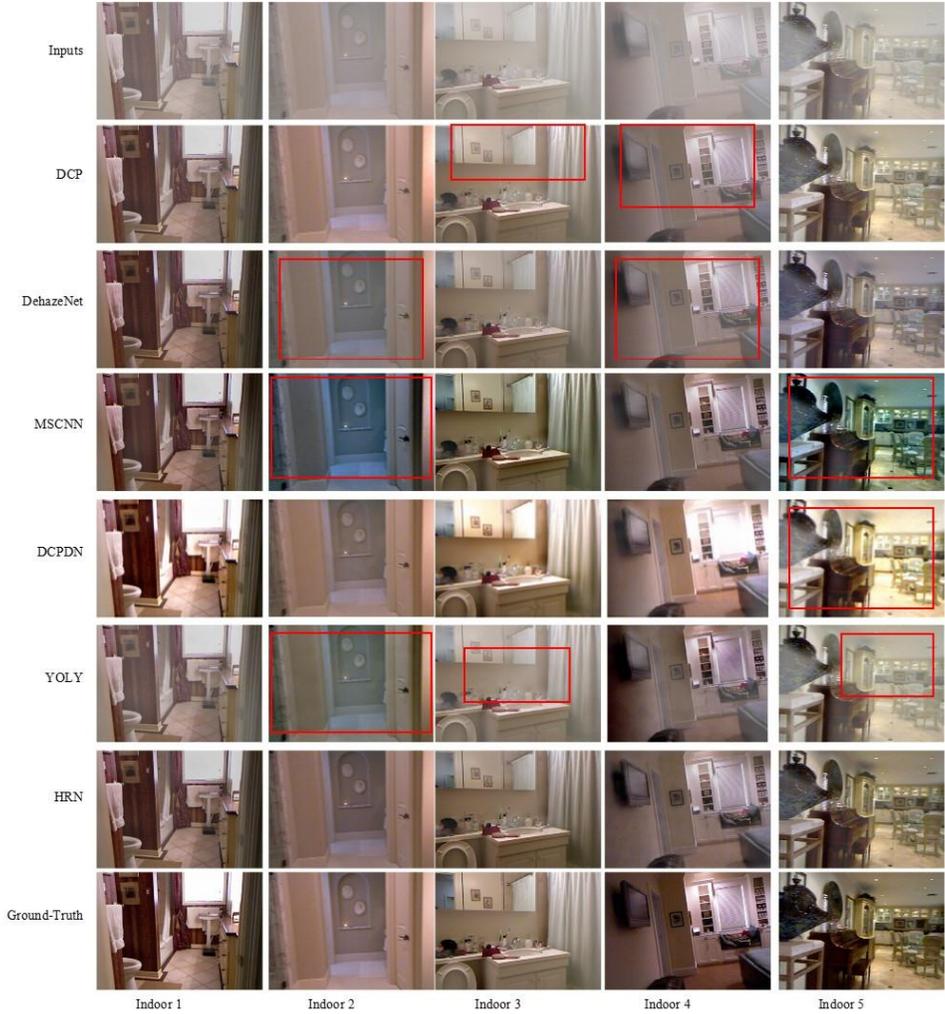
Figure 6. Comparison of different dehazing methods on the synthetic indoor images. The different dehazed results are shown in each row.

from the undesirable colour cast problem. Note that the distant region of Pumpkin image is changed into green, and the sky colour of the City image is distorted to yellow. Similarly, DehazeNet can produce the realistic dehazed results, but some hazes are left in some areas of the dehazing results caused by the inaccurate learning of transmission maps, e.g., the wall of the House image and the sky region of the City image have some remaining hazes. The results of DCPDN have the problem of over-enhancement due to excessive dehazing. The dehazing results of YOLY have the blurred phenomenon, such as the areas in the red boxes of Railway, Pumpkin

Figure 7. Comparison of different dehazing methods on the synthetic outdoor images. The different dehazed results are shown in each row.

and City. Compared with the eight dehazing algorithms, HRN adds the haze relevant information to the nonlinear regression model to remove haze, and it helps to generate more natural and clearer dehazing results, so the colour cast and the most artifacts are suppressed, and there is a similar tone between the original input images and our dehazed results. Furthermore, because the multi-object constraints cost function effectively restrains the parameters of the Haze-Veil model, it avoids outputting the overestimated dark results and obtains the moderately recovered dehazed results.

Figure 8. Comparison of different dehazing methods on the samples of real-world images. The different dehazed results are shown in each row.

4.5 Ablation Experiments

4.5.1 Effectiveness of Haze-Veil Model

To prove that the proposed Haze-Veil could achieve much better dehazed results, we have done a lot of comparison experiments between the traditional two-branch network based on the atmospheric light scattering model and the proposed method, and the results are shown in Figure 9. Specifically, the traditional two-branch network is built based on the traditional atmospheric light scattering model, and is realized by learning V and $t(x)$. To be fair, the base parameters (e.g., the number of training epochs, the number of network layers, the number of feature maps, and the of kernel size of each layer) of the traditional two-branch network are set same as our

proposed network. As observed from Figure 9, with the increasing of the epochs, the average PSNR of proposed method is higher than that of the traditional two-branch network. The results presented here can confirm that the proposed Haze-Veil model indeed helps to improve the dehazed result which is closer to the ground truth.

4.5.2 Effectiveness of Information Interaction Structure

To demonstrate the effectiveness of the information interaction structure (summation operation), five real-world hazy images are selected to conduct the comparison experiments, as shown in Figure 10. Figure 10 a) shows the haze images. Figures 10 b) and 10 d) are the medium transmission maps $t(x)$ and atmospheric veil maps $\boldsymbol{V}(x)$ obtained by the network with the information interaction structure, respectively.
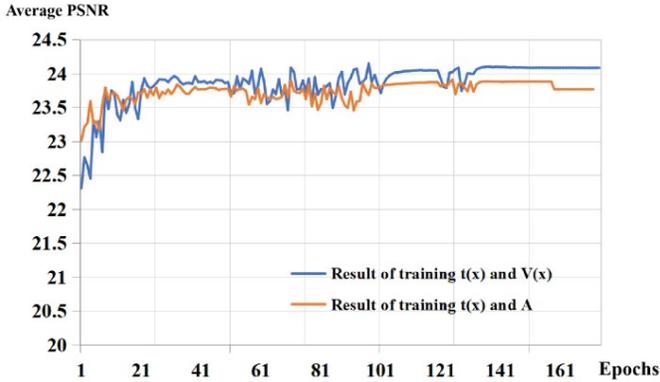
Figure 9. The curve of average PSNR over the training process at synthetic database. The blue curve presents the proposed network built to estimate $t(x)$ and $\boldsymbol{V}(x)$. The orange curve presents the network built to estimate $t(x)$ and $\boldsymbol{A}$.

Figures 10 c) and 10 e) are the medium transmission maps $t(x)$ and atmospheric veil maps $\boldsymbol{V}(x)$ obtained by the network without the information interaction structure, respectively. In Figure 10, the blue and red boxes mark the edges of the regions with large and small depth jumps, respectively. Obviously, a large intensity difference and a clear edge can be observed in the corresponding regions in Figures 10 b) and 10 d). However, the small intensity difference and a blurry edge exist in the corresponding regions in Figures 10 c) and 10 e). Hence, it can be analyzed that the $t(x)$ and $\boldsymbol{V}(x)$ estimated by the network with the information interaction structure can correctly reflect the depth changes, but the $t(x)$ and $\boldsymbol{V}(x)$ estimated by the network without the information interaction structure sometimes failed to accurately address the depth information according to the above analysis.

We further employ the PSNR (peak signal-to-noise ratio) on the synthetic datasets to compare the performance by introducing the information interaction structure in HRN. Comparative experiments prove that the benefit of the summation

a)   b)   c)   d)   e)
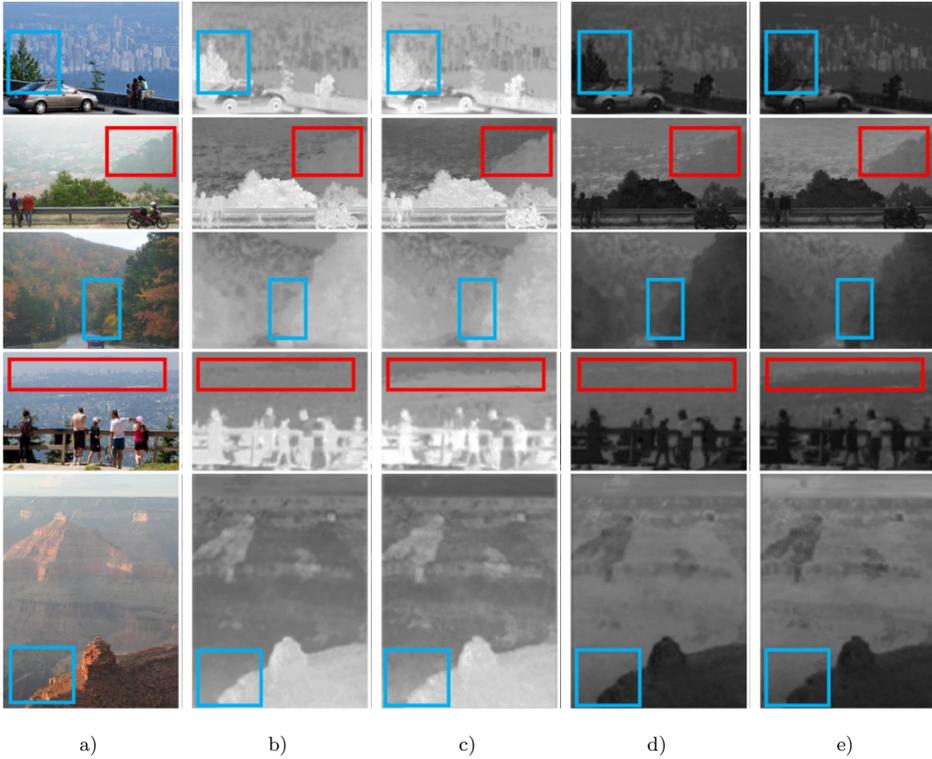
Figure 10. Real-world hazy images with large depth jumps. a) The hazy images. b) The medium transmission map $t(x)$ estimated by the network with information interaction structure. c) The medium transmission map $t(x)$ estimated by the network without information interaction structure. d) The atmospheric veil map $\boldsymbol{V}(x)$ estimated by the network with information interaction structure. e) The atmospheric veil map $\boldsymbol{V}(x)$ estimated by the network without information interaction structure. The regions in blue boxes of the first, third and last rows are with large depth jumps. The regions in red boxes of the second and fourth rows are with small depth jumps.

operation is remarkable. As observed from Figure 11, with the increasing of the epochs, there is a significant improvement in the average PSNR by the summation operation. This finding means that the information interaction structure helps to produce the better dehazed results with higher PSNR values.

4.5.3 Effectiveness of Multi-Object Constraints Cost Function

To prove the advantage of introducing the multi-object constraints cost function in training, the experiments of the average PSNR on synthetic datasets are also conducted. As seen from Figure 12, with the increasing of epochs, the network with
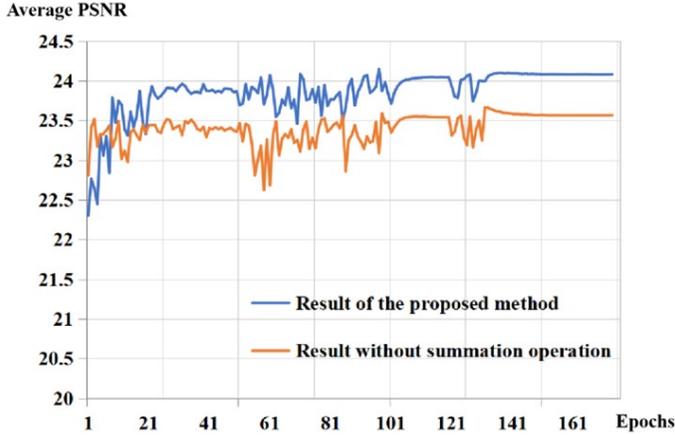
Figure 11. The curve of average PSNR over the training process at synthetic datasets. The blue curve presents the network with summation operation between the two sub-networks. The orange curve presents the network with no summation operation between the two sub-networks.

the multi-object constraints loss in the cost function has a higher average PSNR score than the network only trained by the MSE loss function. In other words, we can obtain the better dehazed image that is closer to the ground-truth. The results also confirm that the defined multi-objective constraint cost function helps to improve the dehazing results and make them closer to the ground-truths.
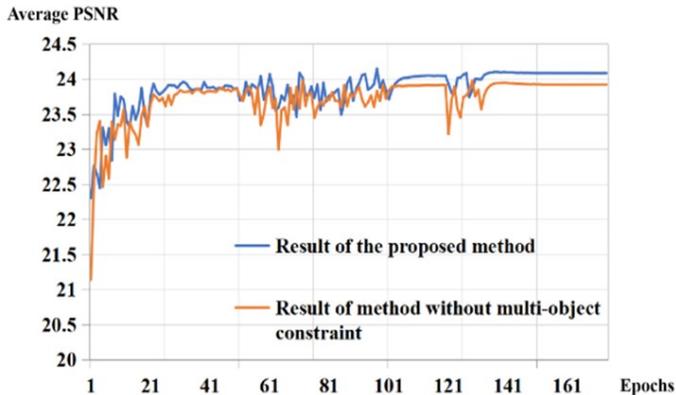
Figure 12. The curve of average PSNR over the training process at synthetic datasets. The blue curve presents the network use the proposed multi-object constraints cost function. The orange curve presents the network only use the Mean Squared Error (MSE) as the cost function.

5 CONCLUSIONS

In this paper, we propose a novel effective Haze-Veil model, which is based on a reformulated atmospheric scattering model to express the degraded mechanism. In this model, the medium transmission and atmospheric veil items are built to remove haze. Then, we design an end-to-end haze-relevant network (HRN) to jointly estimate the two parameters in the Haze-Veil model and directly output the recovered haze-free image. The proposed information interaction structure between the two subnetworks in HRN helps with accurate learning for those parameters. Finally, to generate a constraint relationship between the medium transmission and atmospheric veil, we define a three regularization items-based cost function to efficiently improve the quality of the dehazed image. The experimental results prove that HRN outperforms the state-of-the-art dehazing methods both in synthetic hazy images and real-world hazy images. However, the proposed method is limited to the hazy image with bad light conditions, which will be further researched and optimized in our future work.

Acknowledgements

REFERENCES

[1] Yang, Y.—Wang, C.—Liu, R.—Zhang, L.—Guo, X.—Tao, D.: Self-Augmented Unpaired Image Dehazing via Density and Depth Decomposition. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 2027–2036, doi: 10.1109/CVPR52688.2022.00208.

[2] Wang, Y.—Yan, X.—Wang, F. L.—Xie, H.—Yang, W.—Zhang, X. P.—Qin, J.—Wei, M.: UCL-Dehaze: Toward Real-World Image Dehazing via Unsupervised Contrastive Learning. IEEE Transactions on Image Processing, Vol. 33, 2024, pp. 1361–1374, doi: 10.1109/TIP.2024.3362153.

[3] Tan, R. T.: Visibility in Bad Weather from a Single Image. 2008 IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8, doi: 10.1109/CVPR.2008.4587643.

[4] Tarel, J. P.—Hautière, N.: Fast Visibility Restoration from a Single Color or Gray Level Image. 2009 IEEE 12th International Conference on Computer Vision, 2009, pp. 2201–2208, doi: 10.1109/ICCV.2009.5459251.

[5] Ancuti, C. O.—Ancuti, C.: Single Image Dehazing by Multi-Scale Fusion. IEEE Transactions on Image Processing, Vol. 22, 2013, No. 8, pp. 3271–3282, doi: 10.1109/TIP.2013.2262284.

[6] GALDRAN, A.: Image Dehazing by Artificial Multiple-Exposure Image Fusion. Signal Processing, Vol. 149, 2018, pp. 135–147, doi: 10.1016/j.sigpro.2018.03.008.

[7] XU, Z.—LIU, X.—JI, N.: Fog Removal from Color Images Using Contrast Limited Adaptive Histogram Equalization. 2009 2nd International Congress on Image and Signal Processing, 2009, pp. 1–5, doi: 10.1109/CISP.2009.5301485.

[8] MCDONALD, J. E.: The Saturation Adjustment in Numerical Modelling of Fog. Journal of Atmospheric Sciences, Vol. 20, 1963, No. 5, pp. 476–478, doi: 10.1175/1520-0469(1963)020<0476:TSAINM>2.0.CO;2.

[9] CHAVEZ, P. S.: An Improved Dark-Object Subtraction Technique for Atmospheric Scattering Correction of Multispectral Data. Remote Sensing of Environment, Vol. 24, 1988, No. 3, pp. 459–479, doi: 10.1016/0034-4257(88)90019-3.

[10] FATTAL, R.: Dehazing Using Color-Lines. ACM Transactions on Graphics, Vol. 34, 2015, No. 1, Art. No. 13, doi: 10.1145/2651362.

[11] HE, K.—SUN, J.—TANG, X.: Single Image Haze Removal Using Dark Channel Prior. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, 2011, No. 12, pp. 2341–2353, doi: 10.1109/TPAMI.2010.168.

[12] GIBSON, K. B.—VO, D. T.—NGUYEN, T. Q.: An Investigation of Dehazing Effects on Image and Video Coding. IEEE Transactions on Image Processing, Vol. 21, 2012, No. 2, pp. 662–673, doi: 10.1109/TIP.2011.2166968.

[13] YU, J.—XIAO, C.—LI, D.: Physics-Based Fast Single Image Fog Removal. IEEE 10th International Conference on Signal Processing Proceedings, 2010, pp. 1048–1052, doi: 10.1109/ICOSP.2010.5655901.

[14] MENG, G.—WANG, Y.—DUAN, J.—XIANG, S.—PAN, C.: Efficient Image Dehazing with Boundary Constraint and Contextual Regularization. 2013 IEEE International Conference on Computer Vision, 2013, pp. 617–624, doi: 10.1109/ICCV.2013.82.

[15] NISHINO, K.—KRATZ, L.—LOMBARDI, S.: Bayesian Defogging. International Journal of Computer Vision, Vol. 98, 2012, No. 3, pp. 263–278, doi: 10.1007/s11263-011-0508-1.

[16] BERMAN, D.—TREIBITZ, T.—AVIDAN, S.: Non-Local Image Dehazing. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1674–1682, doi: 10.1109/CVPR.2016.185.

[17] HUANG, S.—WU, D.—YANG, Y.—ZHU, H.: Image Dehazing Based on Robust Sparse Representation. IEEE Access, Vol. 6, 2018, pp. 53907–53917, doi: 10.1109/ACCESS.2018.2871123.

[18] TANG, K.—YANG, J.—WANG, J.: Investigating Haze-Relevant Features in a Learning Framework for Image Dehazing. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2995–3002, doi: 10.1109/CVPR.2014.383.

[19] JIANG, Y.—SUN, C.—ZHAO, Y.—YANG, L.: Fog Density Estimation and Image Defogging Based on Surrogate Modeling for Optical Depth. IEEE Transactions on Image Processing, Vol. 26, 2017, No. 7, pp. 3397–3409, doi: 10.1109/TIP.2017.2700720.

[20] ZHU, Q.—MAI, J.—SHAO, L.: A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior. IEEE Transactions on Image Processing, Vol. 24, 2015, No. 11, pp. 3522–3533, doi: 10.1109/TIP.2015.2446191.

[21] LEVIN, A.—LISCHINSKI, D.—WEISS, Y.: A Closed-Form Solution to Natural Image Matting. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 30, 2008, No. 2, pp. 228–242, doi: 10.1109/TPAMI.2007.1177.

[22] HE, K.—SUN, J.—TANG, X.: Guided Image Filtering. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, 2013, No. 6, pp. 1397–1409, doi: 10.1109/TPAMI.2012.213.

[23] CAI, B.—XU, X.—JIA, K.—QING, C.—TAO, D.: DehazeNet: An End-to-End System for Single Image Haze Removal. IEEE Transactions on Image Processing, Vol. 25, 2016, No. 11, pp. 5187–5198, doi: 10.1109/TIP.2016.2598681.

[24] LI, C.—GUO, J.—PORIKLI, F.—FU, H.—PANG, Y.: A Cascaded Convolutional Neural Network for Single Image Dehazing. IEEE Access, Vol. 6, 2018, pp. 24877–24887, doi: 10.1109/ACCESS.2018.2818882.

[25] REN, W.—PAN, J.—ZHANG, H.—CAO, X.—YANG, M. H.: Single Image Dehazing via Multi-Scale Convolutional Neural Networks with Holistic Edges. International Journal of Computer Vision, Vol. 128, 2020, No. 1, pp. 240–259, doi: 10.1007/s11263-019-01235-8.

[26] LI, B.—PENG, X.—WANG, Z.—XU, J.—FENG, D.: AOD-Net: All-in-One Dehazing Network. 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 4780–4788, doi: 10.1109/ICCV.2017.511.

[27] ZHANG, H.—SINDAGI, V.—PATEL, V. M.: Joint Transmission Map Estimation and Dehazing Using Deep Networks. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 30, 2020, No. 7, pp. 1975–1986, doi: 10.1109/TCSVT.2019.2912145.

[28] REN, W.—MA, L.—ZHANG, J.—PAN, J.—CAO, X.—LIU, W.—YANG, M. H.: Gated Fusion Network for Single Image Dehazing. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 3253–3261, doi: 10.1109/CVPR.2018.00343.

[29] WANG, A.—WANG, W.—LIU, J.—GU, N.: AIPNet: Image-to-Image Single Image Dehazing with Atmospheric Illumination Prior. IEEE Transactions on Image Processing, Vol. 28, 2019, No. 1, pp. 381–393, doi: 10.1109/TIP.2018.2868567.

[30] CHEN, D.—HE, M.—FAN, Q.—LIAO, J.—ZHANG, L.—HOU, D.—YUAN, L.—HUA, G.: Gated Context Aggregation Network for Image Dehazing and Deraining. 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), 2019, pp. 1375–1383, doi: 10.1109/WACV.2019.00151.

[31] SONG, Y.—LI, J.—WANG, X.—CHEN, X.: Single Image Dehazing Using Ranking Convolutional Neural Network. IEEE Transactions on Multimedia, Vol. 20, 2018, No. 6, pp. 1548–1560, doi: 10.1109/TMM.2017.2771472.

[32] YEH, C. H.—HUANG, C. H.—KANG, L. W.: Multi-Scale Deep Residual Learning-Based Single Image Haze Removal via Image Decomposition. IEEE Transactions on Image Processing, Vol. 29, 2020, pp. 3153–3167, doi: 10.1109/TIP.2019.2957929.

[33] YIN, J. L.—HUANG, Y. C.—CHEN, B. H.—YE, S. Z.: Color Transferred Convolutional Neural Networks for Image Dehazing. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 30, 2020, No. 11, pp. 3957–3967, doi: 10.1109/TCSVT.2019.2917315.

[34] DONG, H.—PAN, J.—XIANG, L.—HU, Z.—ZHANG, X.—WANG, F.—

YANG, M. H.: Multi-Scale Boosted Dehazing Network with Dense Feature Fusion. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 2154–2164, doi: 10.1109/CVPR42600.2020.00223.

[35] SHAO, Y.—LI, L.—REN, W.—GAO, C.—SANG, N.: Domain Adaptation for Image Dehazing. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 2805–2814, doi: 10.1109/CVPR42600.2020.00288.

[36] WANG, W.—WANG, A.—LIU, C.: Variational Single Nighttime Image Haze Removal with a Gray Haze-Line Prior. IEEE Transactions on Image Processing, Vol. 31, 2022, pp. 1349–1363, doi: 10.1109/TIP.2022.3141252.

[37] YANG, Y.—HU, W.—HUANG, S.—WAN, W.—GUAN, J.: Progressive Image Dehazing Network Based on Dual Feature Extraction Modules. International Journal of Machine Learning and Cybernetics, Vol. 14, 2023, No. 6, pp. 2169–2180, doi: 10.1007/s13042-022-01753-x.

[38] ZHANG, H.—PATEL, V. M.: Densely Connected Pyramid Dehazing Network. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 3194–3203, doi: 10.1109/CVPR.2018.00337.

[39] LI, B.—GOU, Y.—GU, S.—LIU, J. Z.—ZHOU, J. T.—PENG, X.: You Only Look Yourself: Unsupervised and Untrained Single Image Dehazing Neural Network. International Journal of Computer Vision, Vol. 129, 2021, No. 5, pp. 1754–1767, doi: 10.1007/s11263-021-01431-5.

[40] LI, R.—PAN, J.—LI, Z.—TANG, J.: Single Image Dehazing via Conditional Generative Adversarial Network. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 8202–8211, doi: 10.1109/CVPR.2018.00856.

[41] WU, H.—QU, Y.—LIN, S.—ZHOU, J.—QIAO, R.—ZHANG, Z.—XIE, Y.—MA, L.: Contrastive Learning for Compact Single Image Dehazing. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 10546–10555, doi: 10.1109/CVPR46437.2021.01041.

[42] ZHAO, S.—ZHANG, L.—SHEN, Y.—ZHOU, Y.: RefineDNet: A Weakly Supervised Refinement Framework for Single Image Dehazing. IEEE Transactions on Image Processing, Vol. 30, 2021, pp. 3391–3404, doi: 10.1109/TIP.2021.3060873.

[43] ZEILER, M. D.—FERGUS, R.: Visualizing and Understanding Convolutional Networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.): Computer Vision – ECCV 2014. Springer, Cham, Lecture Notes in Computer Science, Vol. 8689, 2014, pp. 818–833, doi: 10.1007/978-3-319-10590-1_53.

[44] SILBERMAN, N.—HOIEM, D.—KOHLI, P.—FERGUS, R.: Indoor Segmentation and Support Inference from RGBD Images. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (Eds.): Computer Vision – ECCV 2012. Springer, Berlin, Heidelberg, Lecture Notes in Computer Science, Vol. 7576, 2012, pp. 746–760, doi: 10.1007/978-3-642-33715-4_54.

[45] LI, B.—REN, W.—FU, D.—TAO, D.—FENG, D.—ZENG, W.—WANG, Z.: Benchmarking Single-Image Dehazing and Beyond. IEEE Transactions on Image Processing, Vol. 28, 2019, No. 1, pp. 492–505, doi: 10.1109/TIP.2018.2867951.

[46] LIU, Y.—ZHAO, G.—GONG, B.—LI, Y.—RAJ, R.—GOEL, N.—KESAV, S.—GOTTIMUKKALA, S.—WANG, Z.—REN, W.—TAO, D.: Improved Techniques

for Learning to Dehaze and Beyond: A Collective Study. CoRR, 2018, doi: 10.48550/arXiv.1807.00202.

[47] KINGMA, D. P.—BA, J.: Adam: A Method for Stochastic Optimization. CoRR, 2014, doi: 10.48550/arXiv.1412.6980.

[48] WANG, Z.—BOVIK, A. C.—SHEIKH, H. R.—SIMONCELLI, E. P.: Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Transactions on Image Processing, Vol. 13, 2004, No. 4, pp. 600–612, doi: 10.1109/TIP.2003.819861.

[49] LIANG, Y.—WANG, B.—ZUO, W.—LIU, J.—REN, W.: Self-Supervised Learning and Adaptation for Single Image Dehazing. In: De Raedt, L. (Ed.): Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22). 2022, pp. 1137–1143, doi: 10.24963/ijcai.2022/159.

Shuying HUANG (Member, IEEE) received her Ph.D. degree in computer application technology from the Ocean University of China, Qingdao, China, in 2013. She is currently Professor with the School of Software, Tiangong University, Tianjin, China. Her current research interests include image and signal processing, pattern recognition, and deep learning.

Mingyang REN received her B.Eng. degree in computer science and technology from the Hainan University, Haikou, China, in 2020. She is currently pursuing her M.Eng. degree with the School of Computer Science and Technology, Tiangong University, Tianjin, China. Her current research interests include computer vision and deep learning.

Donglei WU is currently working toward his Ph.D. degree majoring in computer science at the Harbin Institute of Technology, Shenzhen, China. His research interests include distributed learning, neural network compression, and image dehazing. He has published several papers in major journals and international conferences including the TPDS, TCAD, TACO, AAAI and ICCD.

Yong YANG (Senior Member, IEEE) received his Ph.D. degree from the Xi'an Jiaotong University, Xi'an, China, in 2005. From 2009 to 2010, he was a postdoctoral research fellow with the Chonbuk National University, Jeonju, Republic of Korea. He is currently Distinguished Professor with the School of Computer Science and Technology, Tiangong University, Tianjin, China. His current research interests include image fusion, image super-resolution reconstruction, medical image processing and analysis, and pattern recognition. He is Associate Editor of the IEEE Access and Editor of the KSII Transactions on Internet and Information Systems.

Weiguo WAN received his B.Sc. degree in mathematics and applied mathematics from the Jiangxi Normal University, Nanchang, China, in 2014, and his Ph.D. degree in computer science and engineering from Jeonbuk National University, Jeonju, South Korea, in 2020. He is currently Lecturer with the School of Software and Internet of Things Engineering, Jiangxi University of Finance and Economics, Nanchang, China. His current research interests include computer vision, deep learning, face sketch synthesis and recognition, and remote sensing image fusion.

Yingzhi WEI received the B.Eng. degree in mechanical and electronical engineering from the Tiangong University, Tianjin, China, in 2021. He is currently pursuing his M.Eng. degree at the School of Computer Science and Technology, Tiangong University, Tianjin, China. His current research interests include computer vision, deep learning and image deraining.